

#6

Method for navigation within a set of audio documents by means  
of a graphic interface  
and receiver for navigation by said method

5        The invention relates to a method of navigation among sound documents accessible with the aid of an audiovisual receiver, and a reproduction apparatus furnished with a graphics user interface making it possible to navigate.

10        The storage of a large number of sound documents within mass-market equipment is known. Mention may be made of audio compact disc (CD) reading apparatus capable of containing a certain number of CDs, a remote control allowing the user to choose on the one hand the appropriate CD, and on the other hand the appropriate piece in this CD. These apparatuses also possess a programming function making it  
15        possible to define a chaining of the sound pieces. During this programming, the user introduces for each piece, the identifier of the CD and the identifier of the piece in the CD. In order to have a certain melodic continuity, the user must know the pieces in advance and program them so as to produce a certain sound continuity during  
20        reproduction.

      Other means of storing sound contents exist. For example, portable readers (or personal players) have a large-capacity electronic memory making it possible to record hundreds of musical pieces. Among the latter may be mentioned the MP3 LYRA reader produced and  
25        manufactured by the applicant. Some home equipment also has a hard disk of large capacity, 20 gigabytes for example, thereby making it possible to store thousands of sound contents.

      The user having access to a large collection of audio contents (for example songs) encounters difficulties in retrieving a determined  
30        piece from his collection, with a view to listening to it. It is therefore

important to arrange them according to certain criteria and especially to present them so that the user can rapidly retrieve the appropriate piece or pieces. It is known to associate digital data for identifying audio contents, the commonest is the title, but there is also: the producer, the singer, the publisher etc. Other elements called "attributes" make it possible to class a content for example the genre (jazz, vocal, rock, easy music, background music, etc.). On the other hand, certain audio contents accessible to a user do not automatically possess these attributes, for example when the user himself records his musical pieces live. Another way of classifying audio contents is to analyse the sound signals directly. Signal analysis techniques exist which make it possible to calculate values of so-called "low-level" parameters for each audio content. These parameters are for example: the tempo, the energy, the brightness, the envelope, etc. They are determined by analysing the signal either in its digital form, or in its analogue form. A technique of audio content indexation is explained in the article "Speech and Language Technologies for audio indexing and retrieval" published in August 2000 in the IEEE Journal page 1338 to 1353 of Volume 88. The article explains how by analysing the audio signal it is possible to classify the various contents. Other articles describe means of calculating low-level parameters and possible uses, here are some other articles included by reference to the present patent application:

- B. Feiten and S. Gunzel, Automatic indexing of a Sound Database using self-organizing neural networks, Computer Music Journal, 18 (3°, 1994
- Eric Scheirer, Music Listening systems, PhD thesis, MIT Media Laboratory, Apr 2000.

The IEEE - WEIPPL document "Visualizing content based relations in texts" published on 29 January 2001, presents various procedures for viewing collections of textual documents by projection into 2D or 3D spaces, employing conventional algorithms such as principal component

analysis or Kohonen maps. The user of such a viewing procedure is a person wishing to search for documents and not a person who prefers to listen to contents without intervening.

Once the low-level parameters have been determined for each  
5 sound document of the collection, the storage or reproduction apparatus  
can class them groupwise as a function of these parameters. Thus, the  
classical music contents may constitute one group, likewise the jazz  
pieces another group. Patent application PCT/GB01/00681 published on  
23 August 2001 describes a user interface consisting of a graphic  
10 displayed on a screen and controlled by an audiovisual receiver. The  
menu displayed exhibits icons ("classical", "jazz", "chart music", "talk  
back", etc.) selectable by the user, the selection of a document of the  
group activating the reproduction of its sound content.

Such interfaces facilitate the selection of an audio content but do  
15 not allow the automatic chaining of several contents. Such chaining may  
be carried out by programming on condition that the user knows the  
various contents in advance. And even in this case, if the user wants to  
obtain melodious chainings, it is not obvious to him how to do it if he  
does not have an ear for music.

20 International Patent Application WO01/65346 – MIHALCHEON  
describes the presentation of an on-line product catalogue. The products  
appear in the form of icons on the screen and the user can select an icon  
thereby triggering the audio reproduction related to the object chosen.  
Passage from one icon to another is effected through navigation according to  
25 a strategy built by the catalogue provider. This navigation cannot therefore  
take into account objects specific to the user's terminal, or criteria specific to  
the user.

The present invention allows a user to successively reproduce  
audio contents contained in his terminal while maintaining a certain  
30 musical unity or at least a certain logic. Moreover, the graphics interface  
thus defined makes it possible to navigate easily within a large collection

of audio content and to reproduce contents that the user desires, doing so in a competitive and user-friendly fashion.

The subject of the invention is a method of navigation within a collection of sound documents stored in a reproduction apparatus furnished with a display device, comprising

- a step of storage of each sound document of the collection,
- a step of analysis of the documents stored so as to determine audio parameters specific to each document, the method is characterized in that it comprises the following steps:

- positioning of graphics identifiers corresponding to at least part of the sound documents on a graphics page of the display device, the position of each graphics identifier being dependent on the parameters calculated previously for a given document,

- automatic navigation by selecting and by reproducing successively the sound documents according to a strategy taking into account the position of the graphics identifiers of the documents in the graphics page and a geometric characteristic specific to the reproduction apparatus.

In this way, the method proposes a novel concept of navigation within a set of audio contents, the concept being based on the viewing of a graphical representation of the set and a strategy based on a graphics relation uniting certain graphics identifiers. Thus, the user can see on the graphics page the evolution of the chainings of the sound documents available within his terminal. Moreover, the position of the representations of the documents in the graphics page depends on low-level parameters calculated for each document, hence navigation based on the position of the representations affords a certain degree of auditory continuity.

According to a first improvement, the method comprises a step of determination of groups of documents possessing close parameter values. The graphics identifiers associated with the documents of a

group are displayed with a similitude allowing the user to locate the group within the graphics page. Thus, the user can choose a sound document within a clearly determined group of music.

According to another improvement, the method comprises a step  
5 of reception of a command for instigating navigation by specifying the navigation strategy used by the receiver for automatically chaining the reproduction of the documents. Several possible navigation strategies exist, all being representable graphically: traversal of a segment, traversal of a spiral or an open shape, definition of a graphics zone  
10 containing identifiers and random selection within this zone.

According to another improvement, the method comprises a step of displaying the number of documents reproduced according to the determined strategy. According to another improvement, the method  
15 comprises a step of displaying the serial number of the document undergoing reproduction.

The subject of the invention is also a reproduction apparatus comprising a central unit, a means of reception of sound documents, a means of storage of the documents received, a means of introduction of  
20 a user command, a means of analysis of the documents stored so as to determine parameters specific to each document, characterized in that it comprises: a means of display of a graphics page of the graphics identifiers corresponding to part at least of the sound documents stored, the position of the graphics identifiers of each document being  
25 dependent on the previously calculated parameters, and a means of navigation for the automatic chaining of the reproduction of the documents according to a determined strategy taking into account the position of the graphics identifiers of the documents within the graphics page and a geometric characteristic specific to the reproduction  
30 apparatus.

Other characteristics and advantages of the invention will now become apparent with greater details within the framework of the description which follows of exemplary embodiments given by way of illustration and referring to the appended figures which represent:

5           - Figure 1 is a block diagram of a reproduction apparatus for the implementation of the invention,

          - Figure 2 is an array associating for each document of the collection its values of low-value parameters,

          - Figure 3 describes a screen shot presenting the collection of  
10 documents in a two-dimensional space,

          - Figure 4 describes a screen shot showing a so-called segment navigation strategy for automatically chaining the sound documents,

          - Figure 5 describes a screen shot showing a spiral navigation strategy for automatically chaining the sound documents.

15

The manner of operation of a reproduction apparatus 1 such as a multimedia receiver 1 associated with a display device 2 will firstly be described. The receiver comprises a central unit 3 linked to a program memory 12, and an interface 5 for communication with a high bit rate  
20 local digital bus 6 making it possible to receive audio and/or video data at high bit rate. This network is for example an IEEE 1394 network. The receiver can also receive audio and/or video data from a transmission network through a reception antenna associated with a demodulator 4, this network can be of radio or television type. The receiver furthermore  
25 comprises a receiver of infrared signals 7 for receiving the signals from a remote control 8, a memory 9 for storing a database, and audio/video decoding logic 10 for generating the audiovisual signals dispatched to the television screen 2. The remote control 8 is fitted with direction keys  $\uparrow$ ,  $\downarrow$ ,  $\rightarrow$  and  $\leftarrow$  and "OK" and "Select" keys whose function we shall  
30 see later.

The receiver also comprises a circuit 11 for displaying data on the screen, often called an OSD circuit, standing for "On Screen Display". The OSD circuit 11 is a text and graphic generator which makes it possible to display menus, pictograms or other graphics on the screen, and menus presenting the navigation. The OSD circuit is controlled by the Central Unit 3 and a navigator 12. The navigator 12 is advantageously embodied in the form of a program module recorded in a read only memory. It may also be embodied in the form of a specialized circuit of ASIC type for example.

The digital bus 6 and/or the transmission network transmit audio contents to the receiver either in digital form, or in analogue form, the receiver recording them in a memory 9. According to a preferred embodiment, the audio contents are received in digital form, preferably coded according to a compression standard, MP3 for example, and stored in the same form. According to this preferred embodiment, the memory 9 is a large-capacity hard disk, 40 gigabytes for example. The storage of a minute of audio content in MP3 occupying around 1 megabyte, such a disk is capable of recording 666 sound hours of document. The downloading of audio content is a well known technique which need not be explained in the present patent application.

Once a certain number of audio contents have been stored in the memory 9. The user wants to reproduce them and to do so without too many manual interventions, he also wants the contents to follow one another with a similitude so as to maintain a harmonious ambiance. To do this, a software module of the navigator analyses each audio content during its reception and extracts the low-level parameters therefrom. As we indicated in the preamble, numerous signal analysis techniques exist which make it possible to obtain arrays of digital descriptors for these songs. The number of elements of a descriptor is of the order of a few tens.

The array contained in the screen page of Figure 2 presents the values of low-level parameters constituting the descriptors of a certain number of audio documents. The first column of the array presents the title of the audio content, each content is numbered. The subsequent  
 5 columns present the values of low-level parameters associated with the document, such as the mean sound intensity, the tempo, the energy, the zero crossing rate, the brightness, the envelope, the bandwidth, the loudness, the cepstral coefficients, etc.

According to an improvement, the low-level parameters may be  
 10 provided in digital form together with the audio content. When the content is provided by a means of digital transmission and in compressed form, the associated low-level parameters constituting a field attached to the audio content. This solution is particularly advantageous since the calculation of the parameters is performed by the producer or the  
 15 provider of the content and not by the user, and hence it is carried out once only.

Be they downloaded or calculated locally, the descriptors are stored in the memory 9 and then utilized to create groups of documents possessing certain similitudes.

20 According to a first approach, the grouping of the contents into coherent groups (or clusters) may be carried out with the aid of a so-called "clustering" algorithm; for example the k-means algorithm (Mac Queen, "Some Methods for classification and analysis of multivariate observations", Proc Fifth Berkeley Symposium on Math., Stat. and Prob.,  
 25 vol1, pp 281-296, 1967.) The array of descriptors of Figure 2 possesses a new column defining the group in which the content is situated. Group calculation techniques are well known, using the k-means algorithm the number of groups thus produced can easily be controlled.

According to a second approach, the groups are determined by a  
 30 prior choice of classes (for example: mood, dominant instruments,



tempo, etc.) and a ground truth helping to define these classes. They are obtained by applying a learning algorithm to this ground truth.

According to the present exemplary embodiment of the invention, the sound documents accessible from the receiver are represented on a screen by graphics identifiers. The position of these graphics identifiers that is to say, the spatial coordinates within the graphics page, are obtained from the low-level parameters. According to the example of Figure 3, the screen represents a two-dimensional navigation space, a point  $P_i$  constituting a graphics identifier representing a sound document  $S_i$ . The coordinates  $(x_i, y_i)$  of a graphics identifier are obtained by projecting the point  $P_i$  whose coordinates are the values of the low-level descriptors onto a sound sample, onto a space of dimension 2, 3, etc. depending on the type of representation chosen. The projection is determined by principal component analysis or PCA. PCA is in particular described in the Saporta 1990 document, entitled "Probabilités Analyse de données et statistiques [Probabilities data analysis and statistics], published by Technip. This well-known data analysis algorithm seeks to discover a subsystem of axes that is linearly bound to the original and which best "spreads" the samples, these axes tend to cause the original correlated axes to merge. The low-level descriptors being assumed to have perceptual coherence (the sounds are perceptually close if and only if the values of the low-level descriptors are close), and the projection being continuous, the graphics identifiers that are close on the screen correspond to perceptually close sounds. This example in no way excludes the representation of the collection by a space with more than two dimensions.

In a general manner, the coordinates  $\{x_i, y_i, \dots, z_i\}$  of a graphics identifier in a multidimensional space allow the user to mentally picture the type of the associated sound document. Specifically, the positions of the graphics identifiers being calculated as a function of the values of low-level parameters, if two identifiers are graphically distant, the values of their low-level parameters are very different and hence, the type of

the sound content is different. On the other hand, if two identifiers are close, the associated audio contents are likewise close in auditory fashion. Data analysis techniques exist which make it possible to discover the predominant dimensions (or dimension combinations) in a certain set of songs. This can be schematized by indicating that these are dimensions corresponding to the axes according to which the songs are most widely distributed. Advantageously, the navigator can analyse the sound documents and determine principal dimensions corresponding to types of audio contents, and it is then the navigator which chooses the number of dimension of the navigation space.

According to an improvement represented also in Figure 3, the sound documents are grouped according to a "clustering" algorithm, for example the k-means algorithm. The graphics identifiers of the elements of one and the same group possess a common characteristic. According to a preferred embodiment, the colour of a graphics identifier depends on the group to which the document belongs (for example: blue, red and green). A variant described in Figure 3 consists in giving the graphics identifier a particular shape: a circle, a cross or a star. An improvement represented in Figure 3 consists in delimiting the groups with the aid of a contour consisting of a closed curved line. In the example illustrated by Figure 3, the navigator has calculated three groups A, B and C, differentiated its members by three particular shapes, then has represented the contours of each group by a closed curve. The graphics identifiers associated with documents of a group appear clustered together on the screen. Specifically, the distribution of the identifiers on the screen is not very generally uniform, groupings of fairly close identifiers appear in the navigation space, these "nebulae" add visual benefit to navigation which chains together sound documents. Also found are isolated identifiers that a curious user may have an urge to listen to. By graphically distinguishing between groups representing different audio content types, the user "sees" his collection and can

choose an audio content by selecting a graphics identifier from the appropriate group. The selection of an identifier is performed by moving a target consisting of two perpendicular straight lines, the intended object being at the intersection of the straight lines. The user moves the target with the aid of the direction keys of his remote control, or of a "joystick". A window at the bottom of the screen displays the title of the audio content undergoing reproduction, if this content suits the user, the latter presses "OK" and the content is reproduced. If the window contains no title, this signifies that the target is not aimed at any audio content.

In the above paragraph, the selection and the reproduction of a single sound document contained in the collection stored in memory 9 is described. We shall now describe automatic navigation over several audio contents.

To instigate the automatic chaining of several audio contents, the user must previously establish a navigation strategy. The user possesses good knowledge of the content of his collection by virtue of the graphic, the latter being reminiscent in a certain manner of nebulae positioned in space, the idea is to establish a path traversing these groups of elements. The representation of the groups of documents of the same type is not a necessity for navigation, nevertheless it affords an aid to the user who can better image how his audio collection is distributed.

We shall now explain several strategies of automatic navigation that the user can select. The objective is to use the graphics representation as a means of definition for the selection of a strategy and for determining the automatic chaining of the documents. The first strategy is that of the straight line segment. This navigation strategy is shown by the drawing of Figure 4. The user selects a starting document Dd (and hence a starting point) and a finishing document Df (respectively a finishing point), and instigates the navigation. The navigator then displays a segment S between these two graphics identifiers and

calculates the distances of each identifier of the collection with the segment. Then, the navigator reproduces the sound document associated with the starting document, then reproduces one after the other the documents situated at a maximum distance from the segment.

5 An improvement consists in displaying a mark (the smiling head of Figures 4 and 5 for example) moving over the segment from the starting point to the finishing point, and in calculating the distances with the graphics identifiers on the basis of this mark. Through the position of the mark on the segment, the user follows the evolution of the navigation  
10 and can determine the time remaining before reaching the finishing point.

According to this navigation strategy, the user introduces three parameters: the coordinates of a starting point (denoted  $D_d$  in Figure 4), the coordinates of a finishing point (denoted  $D_f$  in Figure 4) and the maximum distance (denoted  $d$  in Figure 4) between a graphics identifier  
15 selected by the navigator and the segment. One way of selecting the graphics identifiers consists in moving an index (a square containing the point to be selected) over the screen with the direction keys, the navigator automatically positions the square on a graphics identifier. As to the third parameter, the user keys in a value between 1 and 99.

20 As the distances of each graphics identifier of the collection with the segment are calculated before the first reproduction, the navigator knows the number of documents which will be reproduced successively and displays it in a graphics window in a corner of the screen. In the window is also displayed the serial number of the sound document  
25 undergoing reproduction.

This segment-based navigation strategy makes it possible to go from one type of sound document to another, the transition from one document to the other being made gently since the documents are graphically close.

30 The spiral strategy is represented by Figure 5, the user selects a starting document  $D_d$  (and hence a starting point), a radius of curvature

R and instigates the navigation. The navigator then displays a spiral whose centre is the starting point and whose distance with respect to this identifier increases by the value R with each revolution. As previously, the navigator calculates the distances of each identifier displayed with the spiral. Then, the navigator reproduces the sound document associated with the starting document, then reproduces one after the other the documents situated at a maximum distance from the spiral. In the same way as previously, the navigator displays a mark which moves along the spiral in tandem with reproductions of documents, and, in a graphics window, the number of documents which will be successively reproduced and also the serial number of the sound document undergoing reproduction. The chaining of the reproductions stops when the navigator no longer finds any documents situated at the maximum distance, this generally corresponding to the fact that the index of the spiral has exited the screen.

If the user has placed the starting point bang in the middle of a group of sound documents, this strategy makes it possible to scan a large part of the group and hence to reproduce the same type of document for a long duration. On the other hand, if the user has placed the starting point at the boundary between two groups of sound documents, then with each revolution of the spiral, the navigator reproduces documents of one group then documents of the other group, thereby making it possible to vary the types of reproductions.

These two strategies make it possible to scan a part of the collection according to a chaining that is well determined and hence identically reproducible if the user introduces the same parameters. A third strategy involves a random aspect. The user selects a starting document Dd (and hence a starting point), and a circle radius, and instigates the navigation. The navigator then displays a circle whose centre is the starting point. Next, the navigator randomly selects the graphics identifiers inside the circumscribed zone and reproduces the

associated documents. Navigation stops when all the documents have been reproduced. As previously, the navigator displays in a graphics window the number of documents which will be successively reproduced and also the serial number of the sound document undergoing reproduction. This third strategy has the same advantages as the second (according to the starting point) with additionally that of preventing the documents from being reproduced always with the same chaining.

When the number of documents accessible from the receiver is very large, it is no longer possible to display representations for each of them on the screen. According to an improvement, the graphics page displays a selection of the representations, the user then introduces a criterion for example the genre of the document such as it is defined in the attributes thereof, or else the date of creation of the document or the date of recording, or else if they are songs the name of the singer.

Although the present invention has been described with reference to the particular embodiments illustrated, it is in no way limited by these embodiments, but merely by the appended claims. It will be noted that changes or modifications may be made by the person skilled in the art.